Cost-Performance Tradeoff in Multi-hop Aggregation for Statistical Inference

Animashree Anandkumar*, Lang Tong*, Ananthram Swami† and Anthony Ephremides‡

*ECE Dept., Cornell University, Ithaca, NY 14853, Email: {aa332@,ltong@ece}.cornell.edu
†Army Research Laboratory, Adelphi, MD 20783, Email: a.swami@ieee.org
‡ EE Dept., University of Maryland College Park, MD 20742, Email: tony@eng.umd.edu

***Abstract*—The problem of distributed fusion for binary hypothesis testing in a multihop network is considered. The sensor measurements are spatially correlated according to a Markov random field (MRF) under both the hypotheses. A fusion scheme for detection involves selection and localized processing of a subset of sensor measurements, fusion of these processed values to form a sufficient statistic, and its delivery to the fusion center. The goal is to find a fusion scheme that achieves optimal linear tradeoff between the total routing costs and the resulting detection error exponent at the fusion center. The Neyman-Pearson error exponent, under a fixed type-I bound, is shown to be the limit of the normalized sum of the Kullback-Leibler distances (KLD) over the maximal cliques of the MRF under some convergence conditions. It is shown that optimal fusion reduces to a *prize-collecting Steiner tree* (PCST) with the approximation factor preserved when the cliques of the MRF are disjoint. The PCST is found over an expanded communication graph with virtual nodes added for each non-trivial maximal clique of the MRF and their KLD assigned as the node penalty.**

***Index Terms*—Detection and Estimation, Error Exponents, In-network Processing and Cost-Performance Analysis.**

I. Introduction

The classical paradigm of layered architecture separates design of routing data from the resulting end application performance. Traditionally, data is routed along shortest paths, according to some measure such as energy consumption, without any fusion at the intermediate nodes. However, the severe resource constraints faced by a sensor network has led to recent advances, e.g., network coding [1] and in-network function computation [2] depart from the classical setup and incorporate a data-centric philosophy to reduce routing costs.

For distributed detection, the data-centric paradigm can be taken a step further; only a *sufficient statistic* of the sensor measurements is needed at the designated fusion center for optimal performance. Hence, significant resource savings can be achieved by delivering only the likelihood ratio, the *minimal* sufficient statistic for detection, without destroying information about the underlying phenomenon. However, in general, in-network processing and delivery of the likelihood ratio is a non-trivial problem. By incorporating a *Markov random field* (MRF) statistical model for sensor measurements, we obtained a tractable form for the likelihood ratio and a minimum-cost fusion scheme for optimal detection in [3].

In this paper, we relax the constraint of optimal detection at the fusion center to obtain further savings in resource consumption. Instead, we aim to achieve tradeoffs between resource consumption and the end detection performance by selecting only a subset of "useful" sensor measurements for routing and fusion. Many questions arise with regard to this tradeoff. Can a tractable form of the sufficient statistic and the detection performance for any subset of measurements be obtained? If so, does it have a decentralized form, thereby making the selection process tractable? What is the complexity of the fusion scheme that achieves this optimal tradeoff? Are there efficient approximations to the optimal solution?

A. Related Work

In-network processing algorithms for computing certain aggregate functions are considered in [2]. However, they are applicable only for detection of conditionally-independent data. An overview of routing for detection can be found in [4]. The works in [5] and [6], [7] consider energy-efficient detection for conditionally independent data and one-dimensional Gauss-Markov process, respectively. In [8], we considered the problem of optimal sensor density in an energy-constrained random network, where nodes are placed according to Poisson or uniform distribution. However, in this paper, we assume that the node placement is arbitrary.

The use of Markov random fields in sensor networks is relatively new. The MRF model can capture the full range of correlation between the sensor measurements by varying the edges in a graph, called the *dependency graph*. In [3], we considered the optimal fusion scheme for a MRF model that minimizes the sum routing costs and delivers the likelihood ratio to the fusion scheme. This involved the use of the MRF model to derive a structured form of the likelihood ratio. We describe this problem in detail in the book chapter [9].

B. Our Approach and Contributions

The log-likelihood ratio (LLR) is a minimal sufficient statistic for detection. Hence, maximum reduction in data is achieved by in-network processing of the LLR. The Markov random field (MRF) leads to a succinct form of the LLR. In [3], we proved that the minimum cost fusion scheme that

This work was supported in part through the collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011 and by the National Science Foundation under Contract CNS-0435190. The third author was partially supported by the DARPA ITMANET program. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2008		**00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Cost-Performance Tradeoff in Multi-hop Aggregation for Statistical Inference	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **ECE Department,Cornell University, ,Ithaca,NY,14853**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

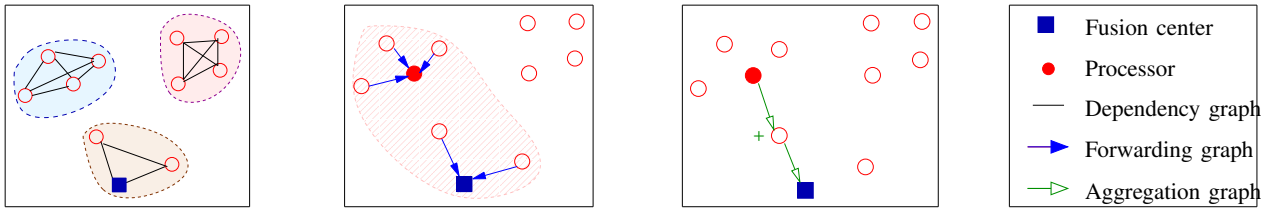
This work was supported in part through the collaborative participation in the Communications and Networks Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011 and by the National Science Foundation under Contract CNS- 0435190. The third author was partially supported by the DARPA ITMANET program. The U. S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

14. ABSTRACT

The problem of distributed fusion for binary hypothesis testing in a multihop network is considered. The sensor measurements are spatially correlated according to a Markov random field (MRF) under both the hypotheses. A fusion scheme for detection involves selection and localized processing of a subset of sensor measurements, fusion of these processed values to form a sufficient statistic, and its delivery to the fusion center. The goal is to find a fusion scheme that achieves optimal linear tradeoff between the total routing costs and the resulting detection error exponent at the fusion center. The Neyman-Pearson error exponent, under a fixed type-I bound, is shown to be the limit of the normalized sum of the Kullback-Leibler distances (KLD) over the maximal cliques of the MRF under some convergence conditions. It is shown that optimal fusion reduces to a prizecollecting Steiner tree (PCST) with the approximation factor preserved when the cliques of the MRF are disjoint. The PCST is found over an expanded communication graph with virtual nodes added for each non-trivial maximal clique of the MRF and their KLD assigned as the node penalty.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Public Release**	**5**	



(a) Maximal cliques of dependency graph (b) Forwarding subgraph computes selected clique potentials (c) Aggregation subgraph to add the potentials (d) Legend

Fig. 1. Schematic of the statistical dependency structure of Markov random field (MRF) and stages of aggregation for detection. The likelihood function is a sum over cliques of MRF. The set of all communication links used during the entire fusion process is the fusion digraph and ensures that the likelihood ratio reaches the fusion center. Its forwarding and aggregation subgraphs transport raw data and aggregated values from selected sensors.

aggregates data from all the sensors, computes the likelihood ratio and delivers it to the fusion center is given by the *Steiner tree* on an expanded communication graph.

In this paper, we consider data fusion over any subset of measurements such that optimal linear tradeoff between the total routing costs and detection error exponent is achieved. First, we obtain a decentralized performance measure over each maximal clique of the MRF. Second, we consider a special class of MRFs and node-selection policies, for which the tradeoff is tractable. Third, we show that optimal tradeoff has a *prize-collecting Steiner tree* (PCST) reduction with the *approximation-factor* preserved.

We first show that under some technical conditions of convergence, the Neyman-Pearson type-II error exponent under a fixed type-I bound is given by the sum of the KLD over the maximal cliques of the MRF. Hence, the KLD of each MRF clique serves as a decentralized performance measure. Since we need to compute and deliver the marginal likelihood ratio of the selected node set to the fusion center, a tractable form is needed. In general for a given subset, computing the marginal likelihood ratio from the joint likelihood ratio is NP-hard. When the MRF contains only disjoint cliques and our policy selects all or none of the nodes of each clique, the marginal likelihood ratio retains its succinct form as product of functions over the selected cliques.

We now design a fusion scheme that can deliver the marginal likelihood ratio of the selected node set to the fusion center and achieve the optimal tradeoff. This optimal fusion scheme is shown to be a prize-collecting Steiner tree (PCST) on an expanded graph. The PCST reduction of optimal fusion implies that any approximation algorithm used for PCST can be applied for the optimal-tradeoff problem with same factor. Note that the approximation factor ρ of any polynomial-time algorithm means that its performance is always no worse than ρ times the optimal value.

A few comments on the use of Markov random field model are in order. The MRF model can incorporate any possible correlation between the measurements. In the trivial case of a general field, the dependency graph of the MRF is complete or fully connected. However, many environmental phenomena such as rainfall or temperature data have been found to possess local dependencies [10] and our scheme has substantial resource savings for such scenarios. However, the use of the MRF model raises issues of learning it from training data and its robustness under changing conditions. These are topics of future interest and in this paper, we assume that all the sensors and the fusion center are aware of the MRF model.

II. System Model

We adopt the system model from our previous work in [3] and provide a brief description in the subsequent sections.

A. MRF Statistical Model for Sensor Data

We assume that the sensor measurements are drawn from a Markov random field. Let $\mathbf{Y}_V = [Y_i, i \in V]^T$ denote the random vector of measurements in set V. $\mathbf{Y}_V$ is a MRF with an undirected dependency graph $G = (V, E)$, if $\forall\, i \in V$,

$$Y_i \perp \mathbf{Y}_{V\setminus\{i,\mathcal{N}_u(i)\}} | \mathbf{Y}_{\mathcal{N}_u(i)}, \tag{1}$$

where $\perp$ denotes conditional independence. In words, the value at any node, given the values at its neighbors, is conditionally independent of the rest of the network. The Hammersley-Clifford theorem [11] states that for PDF f

$$-\log f(\mathbf{Y}_V; \Upsilon) = \sum_{c\in\mathcal{C}} \psi_c(\mathbf{Y}_c), \tag{2}$$

where $\mathcal{C}$ is a collection of (maximal) cliques in G, the functions ψ_c, known as *clique potentials*, are non-negative and the tuple $\Upsilon = \{G, \mathcal{C}, \psi\}$ specifies the MRF. From (2), the complexity of the likelihood function is vastly reduced for sparse dependency graphs and is a product of components, each of which depends on a small set of variables.

B. Detection of Markov Random Field

For binary hypothesis testing with null hypothesis $\mathcal{H}_0$ and alternative hypothesis $\mathcal{H}_1$, we assume that the measurement are drawn from distinct Markov random fields,

$$\mathcal{H}_0 : \Upsilon_0 = \{G_0(V), \mathcal{C}_0, \psi_0\} \text{ vs. } \mathcal{H}_1 : \Upsilon_1 = \{G_1(V), \mathcal{C}_1, \psi_1\}. \tag{3}$$

Let $f(\mathbf{Y}_V;\mathcal{H}_j)$ be the PDF of the measurements $\mathbf{Y}_V$ of sensors in set V under hypothesis $\mathcal{H}_j$. The optimal decision rule is a threshold test based on the log-likelihood ratio (LLR),

$$\text{LLR}(\mathbf{Y}_V) := \log \frac{f(\mathbf{Y}_V;\mathcal{H}_0)}{f(\mathbf{Y}_V;\mathcal{H}_1)}. \quad (4)$$

In [3], it is shown that the LLR has a succinct form

$$\text{LLR}(\mathbf{Y}_V;\Upsilon) = \sum_{c\in\mathcal{C}} \phi_c(\mathbf{Y}_c), \quad (5)$$

where $\Upsilon = \{G_\Upsilon(V), \mathcal{C}, \phi\}$ is the "effective" MRF with dependency graph $G_\Upsilon = G_0 \cup G_1$, clique set $\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1$, with only maximal cliques retained, and potential functions ϕ_c. In this paper, we limit to the scenario where $|\mathcal{C}|$ is polynomial in $|V|$ and the cliques in $\mathcal{C}$ are also the connected components of the dependency graph G(V). In other words, the cliques of the effective MRF are disjoint

$$c_1 \cap c_2 = \emptyset, \quad \forall c_1, c_2 \in \mathcal{C}. \quad (6)$$

One of the scenarios satisfying (6) is when the measurements are independent conditioned on either hypothesis. Another scenario is when there is low-resolution information about the measurement dependencies and sensors are placed in widely-spaced clusters. Here, each cluster is fully connected and there is no edge between different clusters. We can ensure any dependency graph to possess disjoint cliques by adding edges between any two nodes in a connected component and defining new potential functions. However, it is not the sparsest representation for that model.

C. Network and Cost Model

The network is connected via a communication graph containing a set of feasible bidirectional communication links. Note that this communication graph is different from the dependency graph of the MRF. We consider the unicast mode of routing, where a packet from a node is routed to a single destination. The routing of a real number is represented by a packet and quantization error is ignored in our formulation. A symmetric routing cost function is assumed, denoted by $C_{i,j} > 0$. For a set of communication links G, let $\text{C}(G)$ denote the total cost routing using its links.

III. Cost-Performance Tradeoff

We consider the Neyman-Pearson (NP) detection, where for a fixed false-alarm probability, the detector is optimal in terms of the mis-detection probability P_M. We are interested in selecting the measurements from a subset of sensors $V_s \subset V$, processing them using the set of communication links G and delivering them to the designated fusion center. The goal is to achieve optimal linear tradeoff between the total routing cost $\text{C}(G(V_s))$ and the resulting detection performance at the fusion center. Formally, the optimization is

$$\min_{V_s\subset V} \Big[\text{C}(G(V_s)) + \mu \log \frac{P_M(V_s)}{P_M(V)}\Big], \quad \mu > 0, \quad (7)$$

where $\log \frac{P_M(V_s)}{P_M(V)}$ represents the fraction of detection performance we can achieve by selecting the node subset V_s and μ is the tradeoff factor chosen based on the relative importance of the two parameters. When μ is sufficiently large, (7) reduces to minimum cost fusion, considered in [3], where optimal detection is required and hence, all the nodes are selected.

A. Detection Error Exponent

In general, the mis-detection probability $P_M(V_s)$ for an arbitrary node set V_s does not have a closed-form expression and hence, cannot serve as a tractable measure of performance in (7). We focus on the large-network scenario, where $P_M(V)$ decays exponentially with the sample size $|V|$ and we have the NP error exponent

$$\mathcal{D} := -\lim_{|V|\to\infty} \frac{1}{|V|} \log P_M(V). \quad (8)$$

Hence, a large exponent implies faster decay of error probability with increasing sample size. In the lemma below, we now exploit the graphical structure of the MRF to obtain the exponent $\mathcal{D}$ as the limit of a succinct form.

Lemma 1 (NP Error Exponent for MRF): For the NP-detection of (3), under a fixed type-I error bound, when the sequence of normalized LLR random variables $\frac{1}{|V|}\text{LLR}(\mathbf{Y}_V)$ are uniformly integrable and converge in probability under the null hypothesis $\mathcal{H}_0$, the type-II error exponent in (8) is

$$\mathcal{D} = \text{plim}_{|V|\to\infty} \frac{1}{|V|} \sum_{c\in\mathcal{C}} D(\mathbf{f}_0^c || \mathbf{f}_1^c), \quad (9)$$

where plim denotes convergence in probability, $\mathbf{f}_i^c := f(\mathbf{Y}_c;\mathcal{H}_i)$ is the PDF of measurements $\mathbf{Y}_c$ in clique c under $\mathcal{H}_i$, $\mathcal{C}$ is the MRF clique set, and $D(\cdot||\cdot)$ is the Kullback-Leibler distance (KLD).

Proof: When the sequence of normalized LLR converges in probability under null hypothesis[1] , the NP type-II error exponent under a fixed type-I error bound is [13, Theorem 1]

$$\mathcal{D} = \text{plim}_{|V|\to\infty} \frac{1}{|V|}\text{LLR}(\mathbf{Y}_V), \quad \mathbf{Y}_V \sim \mathcal{H}_0, \quad (10)$$

$$= \text{plim}_{|V|\to\infty} \frac{1}{|V|}\mathbb{E}[\text{LLR}(\mathbf{Y}_V);\mathcal{H}_0], \quad (11)$$

where plim denotes convergence in probability. The reduction from (10) to (11) holds when the sequence of the normalized LLR variables is uniformly integrable [12, (16.21)]. Note that

$$\mathbb{E}[\text{LLR}(\mathbf{Y}_V);\mathcal{H}_0] := D(\mathbf{f}_0^V || \mathbf{f}_1^V). \quad (12)$$

[1]Random variables X_n converge in probability to X, if $\lim_n \mathbb{P}[|X_n - X| \geq \epsilon] = 0$, for each positive ϵ. [12, p. 268].

Using the form of LLR for a MRF in (5),

$$D(\mathbf{f}_0^V||\mathbf{f}_1^V) = \sum_{c\in\mathcal{C}} D(\mathbf{f}_0^c||\mathbf{f}_1^c). \tag{13}$$

□

Hence, as the number of nodes goes to infinity, the normalized sum of the clique KLDs tends to the error exponent $\mathcal{D}$ in (15). For the special case when $\mathbf{Y}_V$ are i.i.d. conditioned under each hypothesis, Lemma 1 reduces to the Stein's lemma [14, Theorem 12.8.1] and the limit in (9) to the single-letter KLD. Another case where the exponent $\mathcal{D}$ exists and has a closed form is when the effective dependency graph is the Euclidean nearest neighbor graph, the nodes are placed according to Poisson or uniform distribution and the correlation is a well-defined function of inter-node distance [15].

We now make the asymptotic approximation of misdetection probability for large networks

$$-\log P_M(V) \approx |V|\mathcal{D}. \tag{14}$$

From (9), we have

$$-\log P_M(V) \approx \sum_{c\in\mathcal{C}} D(\mathbf{f}_0^c||\mathbf{f}_1^c), \tag{15}$$

From the disjoint clique property, for any large subset V_s spanning a set of cliques $\mathcal{C}_s$, we have

$$-\log P_M(V_s) \approx \sum_{c\in\mathcal{C}_s} D(\mathbf{f}_0^c||\mathbf{f}_1^c). \tag{16}$$

Hence, the fraction in (7) can be approximated as

$$\log \frac{P_M(V_s)}{P_M(V)} \approx \sum_{c\in\mathcal{C}\setminus\mathcal{C}_s} D(\mathbf{f}_0^c||\mathbf{f}_1^c). \tag{17}$$

This can also be viewed as the *penalty* or the foregone improvement in error exponent due to non-selection of the cliques in $\mathcal{C}\setminus\mathcal{C}_s$. Therefore, the KLD of each clique c of the MRF $D(\mathbf{f}_0^c||\mathbf{f}_1^c)$ serves as a decentralized metric that contributes to the end detection performance. We assume that $D(\mathbf{f}_0^c||\mathbf{f}_1^c) > 0$ for all cliques; otherwise, the clique can be trivially ruled out for selection.

B. Optimal Tradeoff with Localized Processing

Since the log-likelihood ratio (LLR) is the minimal sufficient statistic, maximum savings in routing costs can be achieved by computing the marginal LLR of the selected node set $V_s \subset V$ in (7). Although the LLR over the complete set of sensors V has a succinct form, in terms of the clique potentials in (5), computing the marginal LLR over a general subset V_s is NP-hard. Hence, we limit to the class of policies where we choose all or none of the vertices v_i of each clique. In other words, sensor selection in (7) is limited to the selection of cliques $\mathcal{C}_s \subset \mathcal{C}$,

$$V_s = \{i : i \in V, i \subset \mathcal{C}_s\}, \quad \forall \mathcal{C}_s \subset \mathcal{C}. \tag{18}$$

For the special case of disjoint cliques in (6), the marginal LLR for this node selection policy is still a sum of selected clique potentials,

$$\text{LLR}(\mathbf{Y}_{V_s};\Upsilon) = \sum_{c\in\mathcal{C}_s} \phi_c(\mathbf{Y}_c), \quad \forall \mathcal{C}_s \subset \mathcal{C}. \tag{19}$$

In order to compute a clique potential function ϕ_c in (19), access to measurements of all the clique members are needed. Therefore, if selected for computation, a clique c is assigned a unique computation site, known as its *processor*, denoted by $Proc(c)$. We assume that the clique potential functions are processed "locally", at one of its members, i.e., $Proc(c) \subset c$, $\forall c \in \mathcal{C}_s$ if selected, otherwise, no processor is assigned.

The set of communication links G used by any fusion scheme fall into two categories, viz., those transporting raw measurements to the processor to compute the specified potential function, known as the *forwarding subgraph* $\text{FG}(G)$ and the set of links that transport/aggregate these processed values, known as the *aggregation subgraph* $\text{AG}(G)$. The tuple consisting of the forwarding and the aggregation subgraphs $\{\text{FG}(G), \text{AG}(G)\}$ of a fusion scheme is known as the *fusion digraph*. A schematic of a fusion scheme is shown in Fig.1. Hence, given a MRF $\Upsilon = \{G_\Upsilon, \mathcal{C}, \phi\}$, a fusion scheme is specified by the tuple $\Pi{:=}\{Proc, \text{FG}, \text{AG}, \mathcal{C}_\Pi\}$, with selected clique set $\mathcal{C}_\Pi$, processor assignment mapping $Proc$ and a fusion digraph $\{\text{FG}, \text{AG}\}$. Let $AggVal(i;\Pi)$ be the value at node i at the end of fusion. Formally, the constraints on a fusion scheme Π are

1) the measurements from a node set V_Π spanning a set of cliques $\mathcal{C}_\Pi \subset \mathcal{C}$ is selected for fusion,
2) the marginal LLR of the selected node set V_Π is delivered to the fusion center v_0,

$$AggVal(v_0) = \text{LLR}(\mathbf{Y}_{V_\Pi};\Upsilon) = \sum_{c\in\mathcal{C}_\Pi} \phi_c(\mathbf{Y}_c), \tag{20}$$

3) processor assignment is local for selected cliques,

$$Proc(c) = \begin{cases} i \subset c, i \in V, & \text{if } c \in \mathcal{C}_\Pi, \quad (21a) \\ \emptyset & \text{o.w.} \quad (21b) \end{cases}$$

The fusion scheme achieving optimal tradeoff is given by

$$\Pi^* = \arg\min_{\Pi} [\mathbf{C}(G) + \mu \sum_{c\notin\mathcal{C}_\Pi} D(\mathbf{f}_0^c||\mathbf{f}_1^c)], \tag{22}$$

over all the fusion schemes Π satisfying the above constraints.

IV. Prize-Collecting Steiner-Tree Reduction

In this section, we show that the fusion scheme achieving optimal tradeoff in (22) has a *prize-collecting Steiner-tree reduction* with the approximation factor preserved. We specify the graph transformations required for such a reduction.

The prize-collecting Steiner tree (PCST) is the sub-tree rooted at a specified vertex that minimizes the sum of edge costs in the tree plus the penalties of the vertices not spanned by the tree. Formally, given an undirected graph $G = (V, E)$, a root vertex $v_0 \in V$, nonnegative edge costs $c_e \geq 0, e \in E$, and nonnegative vertex penalties $\pi_i > 0, i \in V$, the PCST is the tree $T^* = (V^*, E^*)$ rooted at v_0,

$$T^* := \arg \min_{T=(V',E')} \{ \sum_{e \in E'} c_e + \sum_{i \notin V'} \pi_i \}. \qquad (23)$$

The penalty associated with a node represents the foregone "profits" for not selecting it and in our setup, the profits represent the end detection error exponent. In general, finding the PCST is NP-hard. The approximation factor ρ of any polynomial-time algorithm guarantees that its performance is no worse than ρ times the optimal value. An approximation algorithm for the PCST, referred to as the *Goemans-Williamson* (GW) algorithm, was proposed in [16]. It has an approximation ratio of $2 - (n-1)^{-1}$ and runs in $O(n^2 \log n)$ time, for a n-node network.

We now provide an algorithm for PCST reduction (AggPCST) of optimal fusion on the lines of the Steiner tree reduction algorithm (AggApprox) [3, Fig.3] for minimum cost fusion. AggApprox incorporates an expansion $Map(G_t)$ [3, Fig.4] on the metric communication graph G_t. It involves adding new virtual nodes called the *clique-representative* nodes corresponding to each non-trivial clique (size greater than one) and connecting it to all its corresponding clique members. The edge cost from a representative node to a clique member incorporates the raw-data routing costs to compute a clique potential function. After constructing the Steiner tree (or its approximation) on the transformed graph $Map(G_t)$, it is mapped to a feasible fusion scheme using the operation $RevMap$ [3, Fig.5]. For the PCST reduction (AggPCST), the following changes are made in AggApprox:

- In $Map(G_t)$, every virtual node representing a clique $c \in \mathcal{C}$ is assigned a penalty $\mu D(\mathbf{f}_0^c || \mathbf{f}_1^c)$, and penalty is zero for all other nodes,
- instead of the Steiner-tree approximation algorithm, the GW-algorithm for PCST is used,
- fusion center is assigned as the root v_0.

In the theorem below, we formally state the PCST reduction.

Theorem 1 (PCST reduction): Given a Markov random field $\Upsilon = \{G_\Upsilon(V), \mathcal{C}, \phi\}$, with dependency graph $G_\Upsilon(V)$, clique set $\mathcal{C}$ with polynomial cardinality and disjoint cliques and potentials ϕ, the AggPCST algorithm outputs a feasible fusion scheme for the optimal cost-performance tradeoff problem in (22) with an approximation factor of $2 - [|V| + |\mathcal{C}'| - 1]^{-1}$, where $\mathcal{C}'$ is the non-trivial clique set (size greater than one).

Proof: The GW-algorithm for PCST is on the expanded graph $Map(G_t)$ with $|V| + |\mathcal{C}'|$ nodes. On lines of [3, Thm. 2], the reduction preserves the approximation factor. □

V. Conclusion

We considered the problem of selection and fusion of sensor measurements such that optimal linear tradeoff between the routing costs and the resulting detection performance at the fusion center is achieved. We exploited the structure of the Markov random field model of sensor measurements to cast it as a prize-collecting Steiner tree, a well-studied combinatorial optimization problem.

We have ignored many issues to keep the problem tractable. We have not addressed the effect of quantization, a difficult problem even for independent measurements. Also, we have only considered offline and centralized sensor selection. Extension to the case where nodes locally decide to transmit based on their measurement values is of interest.

Acknowledgment

The first author thanks Prof. D.P. Williamson and Y. Sharma for extensive discussions and helpful comments on PCST.

References

[1] R. Koetter and M. Medard, "An algebraic approach to network coding," *IEEE/ACM Tran. on Networking*, vol. 11, no. 5, pp. 782–795, 2003.

[2] A. Giridhar and P. Kumar, "In-network information processing in wireless sensor networks," in *Wireless sensor networks: signal processing & comm.* J. Wiley & Sons, 2007, ch. 3, pp. 43–68.

[3] A. Anandkumar, L. Tong, A. Swami, and A. Ephremides, "Minimum Cost Data Aggregation with Localized Processing for Statistical Inference," in *Proc. of IEEE INFOCOM*, Phoenix, USA, April 2008.

[4] S. Misra, L. Tong, and A. Ephremides, "Application dependent shortest path routing in ad-hoc sensor networks," in *Wireless sensor networks: signal processing & comm.* J. Wiley & Sons, 2007, ch. 11, pp. 277–310.

[5] L. Yu, L. Yuan, G. Qu, and A. Ephremides, "Energy-driven detection scheme with guaranteed accuracy," in *Proc. IPSN*, 2006, pp. 284–291.

[6] Y. Sung, S. Misra, L. Tong, and A. Emphremides, "Cooperative Routing for Signal Detection in Large Sensor Networks," *IEEE J. Select. Area Comm.*, vol. 25, no. 2, pp. 471–483, 2007.

[7] J. Chamberland and V. Veeravalli, "How Dense Should a Sensor Network Be for Detection With Correlated Observations?" *Information Theory, IEEE Transactions on*, vol. 52, no. 11, pp. 5099–5106, 2006.

[8] A. Anandkumar, L. Tong, and A. Swami, "Detection of Gauss-Markov Random Fields under Routing Energy Constraint," in *Proc. of 45-th Allerton Conf. on Communication, Control and Computing*, Monticello, USA, Sept. 2007.

[9] A. Anandkumar, A. Ephremides, L. Tong, and A. Swami, "Minimum cost routing with local processing for distributed statistical inference," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and R. Liu, Eds. IEEE-Wiley, 2008, ch. 25.

[10] N. Cressie, *Statistics for spatial data*. J. Wiley, 1993.

[11] J. Hammersley and P. Clifford, "Markov fields on finite graphs and lattices," *Unpublished manuscript*, 1971.

[12] P.Billingsley, *Probability and Measure*. New York, NY: Wiley Inter-Science, 1995.

[13] P.-N. Chen, "General formulas for the Neyman-Pearson type-II error exponent subject to fixed and exponential type-I error bounds," *IEEE Trans. on Information Theory*, vol. 42, no. 1, pp. 316–323, Jan. 1996.

[14] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[15] A. Anandkumar, L. Tong, and A. Swami, "Detection of Gauss-Markov random fields with nearest-neighbor dependency," *Submitted to IEEE Trans. Information Theory*, Dec. 2006, available on Arxiv.

[16] M. Goemans and D. Williamson, "A General Approximation Technique for Constrained Forest Problems," *SIAM Journal on Computing*, vol. 24, p. 296, 1995.